PATENT ABSTRACTS OF JAPAN

(11)Publication number : 2003-345379
(43)Date of publication of application : 03.12.2003

(51)Int.Cl.
G10L 15/00
G10L 15/22
G10L 15/28
H04N 5/278

(21)Application number : 2003-068440
(71)Applicant : JAPAN SCIENCE & TECHNOLOGY CORP
BUG INC
IFUKUBE TATSU
(22)Date of filing : 13.03.2003
(72)Inventor : IFUKUBE TATSU

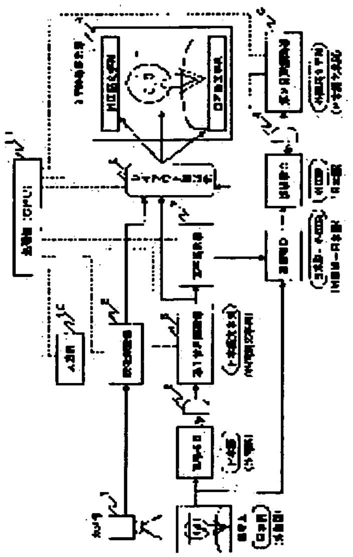
(30)Priority
Priority number : 2002077773 Priority date : 20.03.2002 Priority country : JP

BEST AVAILABLE COPY

(54) AUDIO VIDEO CONVERSION APPARATUS AND METHOD, AND AUDIO VIDEO CONVERSION PROGRAM

(57)Abstract:
PROBLEM TO BE SOLVED: To allow speech of a speaker to be easily understood by recognizing voice of a repeating person repeating the speech of the speaker and displaying a video of the speaker together with characters after a delay.

SOLUTION: A video delay unit 2 outputs delayed video data of a video inputted to a camera 1. A first speech recognition part 5 recognizes contents of a first language of a first repeating person inputted to a first speech input unit 3 and converts it to visible language data. A second speech recognition unit 6 recognizes contents of a second language of a second repeating person inputted to a second speech input unit 4 and converts it to second visible language data. A layout setting unit 8 receives the first and second language data from the first and second speech recognition unit 5 and 6 and delayed video data from the video delay unit 2 and sets a display layout of these data and creates a display video and displays it on a character video display unit 9.



LEGAL STATUS
[Date of request for examination] 26.05.2005
[Date of sending the examiner's decision of rejection]

[Kind of final disposal of application other than the examiner's decision of rejection or application converted registration]

[Date of final disposal for application]

[Patent number]

[Date of registration]

[Number of appeal against examiner's decision of rejection]

[Date of requesting appeal against examiner's decision of rejection]

[Date of extinction of right]

Copyright (C); 1998,2003 Japan Patent Office

* NOTICES *

JPO and NCIPI are not responsible for any damages caused by the use of this translation.

1.This document has been translated by computer. So the translation may not reflect the original precisely.
2.**** shows the word which can not be translated.
3.In the drawings, any words are not translated.

DETAILED DESCRIPTION

[Detailed Description of the Invention]
[0001]
[Field of the Invention] This invention relates to a voice image inverter and an approach, and a voice image conversion program.
[0002]
[Description of the Prior Art] Conventionally, as an exchange means of a meeting by which a hearing-impaired person can participate, there are title broadcast and an epitome note, for example. On the other hand, at present, before using the voice automatic-recognition technique by the computer, it reads out the word and text of some [a user's voice] beforehand, inputs them into a voice recognition unit, and takes how to register the description of a user's voice into a dictionary. Thus, even if it registers a speaker's voice and restricts subject, the highest recognition rate is about at most 95%. Although this invention person has not discovered the report of the paper which conflicts with this invention, in case NHK attaches a title to a broadcast image, the speech recognition method by the repetition person is taken in. Moreover, the report "the nonlinear alphabetic character revitalization software (mospy) by speech recognition is released newly" is announced for Daikin Industries, LTD. by Press Releases (January 20, 2003). This is software which repeats an image and voice, repeating a halt and playback and carries out literation through a voice recognition unit.
[0003]
[Problem(s) to be Solved by the Invention] However, about such conventional title broadcast and a conventional epitome note, it had a big obstruction towards spread that it is not different language correspondence, that making a title or making an epitome take skill, that there are few the experts, etc. On the other hand, about the usual voice automatic-recognition technique, the speech recognition of the unspecified speaker now correctly recognized also in whose voice has a very low precision, and the case where it cannot be used is assumed under an environment with many noises. Moreover, the audio recognition time will take about 1 second, and a pan will take 2 to 3 seconds through a translator. Big time difference arises on the expression of the character string which is the result of carrying out speech recognition, and a speaker etc., therefore it becomes impossible therefore, to use vision data, such as sign language, for a motion and expression of a speaker's lip, and a pan at an understanding of meaning of a passage. Furthermore, in the case of Japanese, since there is a kanji of many homophenes, if meaning of a passage cannot be presumed from a context, it will incorrect-change. With the current technique, it is difficult to grasp meaning of a passage artificially, and he has left selection of the kanji to the user of a voice recognition unit. Moreover, with a current speech recognition technique, if a speaker and subject change, a recognition rate will fall just then. It is restricted to a quiet place, and moreover, an operating environment must also use a specific thing and must also always install a microphone in the same location of the month. Thus, it was difficult to use a voice recognition unit for the meeting exchange and interpreter exchange for a hearing-impaired person conventionally. Furthermore, since telecommunication circuits, such as the Internet, were not used for the above-mentioned NHK method and the above-mentioned Daikin Industries, LTD. product, the service which supports a user by the translator who is present in a remote

place or the home ground, and the repetition person was not able to be offered.

[0004] This invention aims at offering a voice image inverter for a hearing-impaired person etc. making easy to understand what the speaker spoke about and an approach, and a voice image conversion program by delaying images, such as a speaker's expression, and displaying on a screen etc. with an alphabetic character while a repetition person changes the voice of an unspecified speaker into self voice and changes it into an alphabetic character through a voice recognition unit in view of the above point. Moreover, in the meeting of an international congress which a hearing-impaired person attends, a bilateral meeting between many countries, etc., etc., a repetition person repeats the voice of a lecturer or a translator, and this invention inputs it into a voice recognition unit, and aims at offering the voice image inverter for the meeting exchange which displayed the character string which it is as a result on the screen with a lecturer's image and an approach, and a voice image conversion program. Furthermore, this invention aims at providing a user with the information by which transmitted voice to the repetition person and literation was carried out from exchange of the meeting where the interpreter of the international congress performed using different-species language and the immediate printing (information compensation) of a meeting, a hearing-impaired person, etc. participate, or a lesson, and a telephone. Moreover, this invention aims at offering the voice image inverter for assisting communication between different language systems of a speaker and a user and an approach, and a voice image conversion program. Moreover, by adding a means to transmit a speaker's voice and image to the translator, the repetition person, and those [correction] who are present in a remote place or the home ground by the telecommunication circuit which communicates using telecommunication circuits, such as the Internet, further, wherever the user of this invention may be in, it aims at enabling it to use this system. This invention aims at that the intervening repetition person and a translator use as home business, and supporting working, when the trouble back tone of difficult being home of going out turns into a repetition person further.

[0005]

[Means for Solving the Problem] The camera which photos a speaker's expression image according to the 1st solution means of this invention, The image delay section which gives the time delay difference beforehand set up to the video signal photoed with said camera, and outputs delay image data, The 1st voice input section into which the contents of the 1st language of the 1st repetition person who repeats the contents of the 1st language about which a speaker speaks are inputted, The 2nd voice input section into which the contents of the 2nd language of the 2nd repetition person who repeats further the contents of the 2nd language of the translator who interpreted the contents of the 1st language about which a speaker speaks are inputted, the contents of the 1st and 2nd language inputted from said 1st and 2nd voice input section, respectively -- recognizing -- the [the 1st and] -- with the 1st and 2nd speech recognition section changed and outputted to 2 visible language data the [the 1st outputted from said 1st and 2nd speech recognition section, and] -- with 2 visible language data The layout setting section which generates the display image which the delay image data of the speaker delayed by said image delay section were inputted [image], and the display condition was set [image] up, and synchronized or synchronized [abbreviation] these data, the output from said layout setting section -- following -- the [the 1st and] -- the alphabetic character graphic display section which displays the display image which synchronized or synchronized [abbreviation] 2 visible language data and delay image data, and said 1st and 2nd speech recognition section -- The voice image inverter equipped with the input section for performing various setup of said image delay section, said layout setting section, or two or more each part and the processing section which controls each part of said 1st and 2nd speech recognition section, said image delay section, said input section, and said layout setting section is offered.

[0006] The camera which photos a speaker's expression image according to the 2nd solution means of this invention, The image delay section which gives the time delay difference beforehand set up to the video signal photoed with said camera, and outputs delay image data, The 1st voice input section into which the contents of the 1st language of the 1st repetition person who repeats the contents of the 1st language about which a speaker or a translator speaks are inputted, and the 1st speech recognition section the contents of the 1st language inputted from said 1st voice input section -- recognizing -- the -- with the 1st speech recognition section changed and outputted to 1 visible language data One visible language data and the delay image data of the speaker delayed by said image delay section are inputted. the [which was outputted from said 1st speech recognition section] -- The layout setting section which generates the display image which the display condition was set [image] up, and synchronized or synchronized [abbreviation] these data, the output from said layout setting section -- following -- the -- the alphabetic character graphic display section which displays the display image which synchronized or synchronized [abbreviation] 1 visible language data and delay image data, and said 1st speech recognition section -- The voice image inverter equipped with the input section for performing various setup of said image delay section, said layout setting section, or two or more each part and the processing section which controls each part of said 1st speech recognition section, said image delay section, said input section, and said layout setting section is offered.

[0007] It is the voice image conversion approach or program for according to the 3rd solution means of this invention, changing a speaker's voice into visible language data, and displaying with a speaker's image data. The processing section According to a setup beforehand defined by the command from the input section, or the proper storage section, the step which performs a setup of the 1st and 2nd speech recognition section and the image delay section, and the processing section According to a setup beforehand defined by the command from the input section, or the proper storage section, the step which sets up the layout setting section, and a camera The step which inputs a speaker's image, and the image delay section The step which performs a proper image processing for the image inputted into the camera delay and if needed according to a setup and control by the processing section, and outputs delay image data, and the 1st voice input section The step which inputs the contents of the 1st language by the 1st repetition person who repeats the contents of the 1st language by the speaker, and the 1st speech recognition section the contents of the 1st language by the 1st repetition person inputted into the 1st voice input section -- recognizing -- the -- the step changed into 1 visible language data and the 2nd voice input section The step which the 2nd repetition person repeats the contents of the 2nd language with which the translator interpreted the contents of the 1st language by the speaker, and inputs the contents of the 2nd repeated language, and the 2nd speech recognition section the contents of the 2nd language by the 2nd repetition person inputted into the 2nd voice input section -- recognizing -- the -- the step changed into 2 visible language data and the layout setting section According to a setup and control by the processing section, the 1st and 2nd language data from the 1st and 2nd speech recognition section and the delay image data from the image delay section are inputted. The step which generates and outputs the display image which the display layout of these data was set [image] up, and synchronized or synchronized [abbreviation] these data by the image processing, and the alphabetic character graphic display section The program for making a computer perform the voice image conversion approach containing the step which displays the display image which synchronized or synchronized [abbreviation] the 1st and 2nd language data and image lag data, and each [these] step according to the output from the layout setting section is offered.

[0008] It is the voice image conversion or the program for according to the 4th solution means of this invention, changing a speaker's voice into visible language data, and displaying with a speaker's image data. The processing section According to a setup beforehand defined by the command from the input section, or the proper storage section, the step which performs a setup of the 1st speech recognition section and the image delay section, and the processing section According to a setup beforehand defined by the command from the input section, or the proper storage section, the step which sets up the layout setting section, and a camera The step which inputs a speaker's image, and the image delay section The step which performs a proper image processing for the image inputted into the camera delay and if needed according to a setup and control by the processing section, and outputs delay image data, and the 1st voice input section The step which inputs the contents of the 1st language by the 1st repetition person who repeats the contents of the 1st language by the speaker or the translator, and the 1st speech recognition section the contents of the 1st language by the 1st repetition person inputted into the 1st voice input section -- recognizing -- the -- the step changed into 1 visible language data and the layout setting section According to a setup and control by the processing section, the 1st language data from the 1st speech recognition section and the delay image data from the image delay section are inputted. The step which generates and outputs the display image which the display layout of these data was set [image] up, and synchronized or synchronized [abbreviation] these data by the image processing, and the alphabetic character graphic display section The program for making a computer perform the voice image conversion approach containing the step which displays the display image which synchronized or synchronized [abbreviation] the 1st language data and image lag data, and each [these] step according to the output from the layout setting section is offered.

[0009] the contents of the 1st language of the 1st repetition person who repeats the contents of the 1st language a speaker says that it is based on the 5th solution means of this invention -- recognizing -- the -- with the 1st speech recognition section changed and outputted to 1 visible language data The 1st recognition equipment which has the 1st input section for performing various setup of said 1st speech recognition section, and the 1st processing section which controls said 1st speech recognition section and said 1st input section, the contents of the 2nd language of the 2nd repetition person who repeats further the contents of the 2nd language of the translator who interpreted the contents of the 1st language about which a speaker speaks -- recognizing -- the -- with the 2nd speech recognition section changed and outputted to 2 visible language data The 2nd recognition equipment which has the 2nd input section for performing various setup of said 2nd speech recognition section, and the 2nd processing section which controls said 2nd speech recognition section and said 2nd input section, The output from said 1st and 2nd recognition equipment is inputted, and it has a display for displaying an alphabetic character and an image. Said display The image delay section which gives the time delay difference beforehand set up to the video signal photoed with the camera, and outputs delay image data, the [the 1st outputted from said 1st and 2nd recognition equipment, and] -- with 2 visible language data The layout setting section which generates the display image which the delay image data of the speaker delayed by said image delay section were inputted [image], and the display condition was set [image] up, and synchronized or synchronized [abbreviation] these data, The 3rd input section for performing various setup of the alphabetic character graphic display section which displays the display image outputted from said layout setting section, said image delay section, and said layout setting section, The voice image inverter which has the 3rd processing section which controls each part of said image delay section, said 3rd input section, and said layout setting section is offered. the contents of the 1st language of the 1st repetition person who repeats the contents of the 1st language a speaker or a translator says that it is based on the 6th solution means of this invention -- recognizing -- the -- with the 1st speech recognition section changed and outputted to 1 visible language data The 1st recognition equipment which has the 1st input section for performing various setup of said 1st speech recognition section, and the 1st processing section which controls said 1st speech recognition section and said 1st input section, The output from said 1st recognition equipment is inputted, and it has a display for displaying an alphabetic character and an image. Said display The image delay section which gives the time delay difference beforehand set up to the video signal photoed with the camera, and outputs delay image data, One visible language data and the delay image data of the speaker delayed by said image delay section are inputted. the [which was outputted from said 1st recognition equipment] -- The layout setting section which generates the display image which the display condition was set [image] up, and synchronized or synchronized [abbreviation] these data, The 3rd input section for performing various setup of the alphabetic character graphic display section which displays the display image outputted from said layout setting section, said image delay section, and said layout setting section, The voice image inverter which has the 3rd processing section which controls each part of said image delay section, said 3rd input section, and said layout setting section is offered. It is the voice image conversion approach for according to the 7th solution means of this invention, changing a speaker's voice into visible language data, and displaying with a speaker's image data.

The step to which the 1st and 2nd processing section and the 3rd processing section perform a setup of the 1st and 2nd speech recognition section and the image delay section according to a setup as which it was beforehand determined by the command from the 1st and 2nd input section and the 3rd input section, or the proper storage section, respectively, According to a setup as which the 3rd processing section was beforehand determined by the command from the 3rd input section, or the proper storage section, the step which sets up the layout setting section, and the image delay section The step which performs a proper image processing for a speaker's image inputted into the camera delay and if needed according to a setup and control by the 3rd processing section, and outputs delay image data, and the 1st speech recognition section the contents of the 1st language by the 1st repetition person who repeats the contents of the 1st language by the speaker -- recognizing -- the -- the step changed into 1 visible language data and the 2nd speech recognition section the contents of the 2nd language by the 2nd repetition person who repeated the contents of the 2nd language with which the translator interpreted the contents of the 1st language by the speaker -- recognizing -- the -- with the step changed into 2 visible language data Two visible language data and the delay image data from the image delay section are inputted. a setup and control according [the layout setting section] to the 3rd processing section -- following -- the [from the 1st and 2nd speech recognition section / the 1st and] -- The step which generates and outputs the display image which the display layout of these data was set [image] up, and synchronized or synchronized [abbreviation] these data by the image processing, and the alphabetic character graphic display section the output from the layout setting section -- following -- the [the 1st and] -- the voice image conversion approach containing the step which displays the display image which synchronized or synchronized [abbreviation] 2 visible language data and image lag data is offered. It is the voice image conversion approach for according to the 8th solution means of this invention, changing a speaker's voice into visible language data, and displaying with a speaker's image data. The 1st and 3rd processing section According to a setup beforehand defined by the command from the 1st and 3rd input section, or the proper storage section, the step which performs a setup of the 1st speech recognition section and the image delay section, and the 3rd processing section, respectively According to a setup beforehand defined by the command from the 3rd input section, or the proper storage section, the step which sets up the layout setting section, and the image delay section The step which performs a proper image processing for a speaker's image inputted into the camera delay and if needed according to a setup and control by the 3rd processing section, and outputs delay image data, and the 1st speech recognition section the contents of the 1st language by the 1st repetition person who repeats the contents of the 1st language by the speaker or the translator -- recognizing -- the -- the step changed into 1 visible language data and the layout setting section One visible language data and the delay image data from the image delay section are inputted. a setup and control by the 3rd processing section -- following -- the [from the 1st speech recognition section] -- The step which generates and outputs the display image which the display layout of these data was set [image] up, and synchronized or synchronized [abbreviation] these data by the image processing, and the alphabetic character graphic display section the output from the layout setting section -- following -- the -- the voice image conversion approach containing the step which displays the display image which synchronized or synchronized [abbreviation] 1 visible language data and image lag data is offered.

[0010]

[Embodiment of the Invention] Hereafter, the gestalt of operation of this invention is explained to a detail using a drawing.

1. Gestalt drawing 1 of the 1st operation is the outline block diagram of the gestalt of operation of the 1st of a voice image inverter. Especially the gestalt of this operation supports the communication in the meeting, the meeting, the lecture, lesson, education, etc. in which two or more language, such as an international congress, a meeting between many countries, and a meeting between two nations, participates. The voice image inverter of the gestalt of this operation is equipped with a camera 1, the image delay section 2, the 1st and 2nd voice input sections 3 and 4, the 1st and 2nd speech recognition sections 5 and 6, the character

representation section 7, the layout setting section 8, the alphabetic character graphic display section 9, the input section 10, and the processing section 11.

[0011] A camera 1 photos Speaker's A expression image. The image delay section 2 gives the time delay difference beforehand set up to the video signal from a camera 1, and outputs delay image data. The image delay section 2 is displayed together with the alphabetic character which had a speaker's expression image recognized, and in order to make it become assistance of a sink's lalognosis, it gives a predetermined image time delay. This image time delay can be suitably changed according to speed, capacity, etc. about which the speech reading capacity of meeting participants, such as a hearing-impaired person, speaker A and a repetition person B, or C and a translator D speaks. Moreover, the image delay section 2 may be made to perform proper image processings, such as zooming, for images, such as Speaker's A expression.

[0012] The 1st voice input section 3 consists of microphones etc., and contents with the voice of the specific 1st repetition person B who caught Speaker's A voice are inputted. On the other hand, Translator D interprets the contents as which Speaker A says the 2nd voice input section 4, and contents with the voice of the specific 2nd repetition person C who caught the translator's D voice are inputted. The repetition persons B and C are the quiet locations prepared in the meeting, are carrying out voice input through the 1st or 2nd voice input sections 3 and 4, such as a tale microphone, and can also solve the effect of ambient noise or a microphone.

[0013] the voice into which the 1st and 2nd speech recognition sections 5 and 6 were inputted from the 1st and 2nd voice input sections 3 and 4, respectively -- recognizing -- the [, such as alphabetic data and ideography data, / the 1st and] -- it changes and outputs to 2 visible language data. In this example, the contents repeated in the 1st language by the 1st repetition person B who heard the 1st language (example: Japanese) which Speaker A speaks are inputted, and the 1st speech recognition section 5 outputs the visible language data (example: Japanese character string) of the 1st language. On the other hand, the translator D who heard the 1st language (example: Japanese) which Speaker A speaks acts as interpreter in the 2nd language (example: foreign languages, such as English), the contents repeated in the 2nd language by the 2nd repetition person C who heard further the 2nd language which Translator D speaks are inputted, and the 2nd speech recognition section 6 outputs the visible language data (example: foreign language character strings, such as English) of the 2nd language.

[0014] The 1st and/or the 2nd speech recognition sections 5 and 6 may enable it to choose both the voice as which the 1st repetition person B repeated voice, and both [either or] as which the 2nd repetition person C repeated Translator's D voice. The 1st and/or the 2nd speech recognition sections 5 and 6 are set up so that a repetition person's voice may be recognized, and you may make it equipped with the selection section which can choose the language database with which the 1st and/or the 2nd repetition persons B and C are registered into the 1st and/or the 2nd voice recognition unit 5 and 6 according to the subject about which Speaker A speaks, or the contents of the meeting.

[0015] Furthermore, you may make it the 1st and/or the 2nd speech recognition sections 5 and 6 equipped with the incorrect conversion probability count section which calculates the probability incorrect-changed in kana-kanji conversion, and the output decision section which opts for a kanji output or a kana alphabetic character output according to the probability calculated in the incorrect conversion probability count section. About kanji processing of the Japanese homophenes, the 1st and/or the 2nd speech recognition sections 5 and 6 calculate the probability of incorrect recognition before speech recognition, and when the probability is high, they can display it in a kana alphabetic character. Moreover, you may make it display the language which is not registered into the 1st and/or the 2nd speech recognition sections 5 and 6 in a kana alphabetic character by decision of the 1st and/or the 2nd repetition persons B and C.

[0016] The character representation section 7 indicates the visible language data of the 1st language outputted by the 1st speech recognition section 5 by visible. the [as which Translator D was displayed by the character representation section 7] -- 1 visible language data are seen and you may make it act as interpreter

[0017] the [the 1st outputted as a result the layout setting section 8 has been recognized to be by the 1st and 2nd speech recognition sections 5 and 6, and] -- 2 visible language data and the delay image data of the speaker A delayed by the image delay section 2 are inputted, and the display condition to the alphabetic character graphic display section 9 is set up. the [the 1st as which the processing section 11 is displayed on the alphabetic character graphic display section 9, and] -- about 2 visible language data (alphabetic data) and delay image data Either or the plurality of a display format of the line count per unit time amount, the number of alphabetic characters per unit time amount, the number of alphabetic characters of 1 end of a road, a color, magnitude, a display position, and others is set up. The layout setting section 8 a setup by the processing section 11 -- responding -- the [the 1st and] -- proper image processings, such as zooming about 2 visible language data and delay image data, are performed, and a display image is generated.

[0018] the [the 1st outputted according to the output the alphabetic character graphic display section 9 was set up and generated by whose layout setting section 8 as a result recognized by the 1st and 2nd speech recognition sections 5 and 6, and] -- it displays combining 2 visible language data and the delay image data of the speaker A delayed by the image delay section 2. The input section 10 performs the data input directions to various setup, a proper database, memory of each part of the 1st and 2nd speech recognition sections 5 and 6, the image delay section 2, and layout setting section 8 grade, etc. The processing section 11 is a small computer and controls each part of the 1st and 2nd speech recognition sections 5 and 6, the image delay section 2, the input section 10, and layout setting section 8 grade.

[0019] The gestalt flow chart of implementation of the 1st of voice transform processing by the processing section is shown in drawing 2 . The processing section 11 performs a setup of the 1st and 2nd speech recognition sections 5 and 6 and the image delay section 2 according to a setup beforehand defined by the command from the input section 10, or the proper storage section (S01). In a setup of the 1st and 2nd speech recognition sections 5 and 6, the threshold of a kanji incorrect recognition rate, the language database to be used are set up, for example. In a setup of the image delay section 2, a setup or selection of the time delay of a speaker image is performed, for example. Furthermore, the processing section 11 sets up the layout setting section 8 according to a setup beforehand defined by the command from the input section 10, or the proper storage section (S03). the [which is displayed on the alphabetic character graphic display section 9 in a setup of the layout setting section 8 / 1st] -- the display condition and layout of 2 visible language data and delay image data are set up. About the number of presentation character strings, a presentation graphic size, a font and a color, the display position of a character string, and delay image data, the magnitude of a speaker image, a display position, etc. are set [data / visible language] up suitably, respectively, for example.

[0020] A camera 1 inputs Speaker's A image (S05). The image delay section 2 performs a proper image processing for the image inputted into the camera 1 delay and if needed according to a setup and control by the processing section 11, and outputs delay image data (S07).

[0021] The 1st voice input section 3 inputs the voice by the 1st repetition person B (S11). the 1st language by the 1st repetition person B by whom the 1st speech recognition section 5 was inputted into the 1st voice input section 3 according to a setup and control by the processing section 11 -- recognizing -- the -- it changes into 1 visible language data (example: Japanese character string) (S13). the [furthermore, / to which the character representation section 7 was outputted from the 1st speech recognition section 5 if needed] -- 1 visible language data are displayed (S15).

[0022] the [as which, as for the 2nd voice input section 4, Translator D was displayed on speaker voice and/or the character representation section 7] -- the 2nd repetition person C repeats the voice interpreted based on 1 visible language data, and the repeated voice is inputted (S17). the 2nd language by the 2nd repetition person C by whom the 2nd speech recognition section 6 was inputted into the 2nd voice input section 4 according to a setup and control by the processing section 11 -- recognizing -- the -- it changes into 2 visible language data (example: foreign language character string) (S19).

[0023] a setup and control according [the layout setting section 8] to the processing section 11 -- following -- the [from the 1st and 2nd speech recognition sections 5 and 6 / the 1st

and] -- delay image data are inputted from 2 visible language data and the image delay section 2, the display layout of these data is set up, the need is accepted, and a display image is generated and outputted by the proper image processing (S21). the alphabetic character graphic display section 9 -- the output from the layout setting section 8 -- following -- the [the 1st and] -- 2 visible language data and the image delay section 2 are displayed suitably (S23).
[0024] The processing section 11 performs return processing to step S01, when there is setting modification (S25). Moreover, when there is no setting modification, when there is no speaker A modification, it moves to the processing after step S03, and on the other hand, when there is speaker A modification, the processing section 11 can end processing (S27), and can perform processing anew.
[0025] 2. Gestalt drawing 3 of the 2nd operation is the outline block diagram of the gestalt of operation of the 2nd of a voice image inverter. Especially the gestalt of this operation supports the communication in a meeting, a meeting, a lecture, a lesson, education, etc., such as a domestic meeting and a meeting between two nations. The voice image inverter of the gestalt of this operation is equipped with a camera 1, the image delay section 2, the 1st and 2nd voice input sections 3 and 4, the 1st speech recognition section 5, the character representation section 7, the layout setting section 8, the alphabetic character graphic display section 9, the input section 10, the processing section 11, and the selection section 20.
[0026] Although it differs in that the 2nd speech recognition section was omitted as compared with the gestalt of the 1st operation, and it had the selection section 20 further, other configurations and actuation are the same. In addition, the 2nd voice input section and the selection section 20 may be omitted further if needed.
[0027] The gestalt flow chart of implementation of the 2nd of voice transform processing by the processing section is shown in drawing 4 . As compared with the gestalt of the 1st operation, it mainly differs in that steps S17-S19 were skipped. Moreover, either of the voice as which the repetition person C repeated the voice of the translator D who interpreted the voice of the repetition person B who repeated a speaker's voice, and a speaker's voice is inputted into the 1st voice input section 3.
[0028] The processing section 11 performs a setup of the 1st speech recognition section 5, the image delay section 2, and the selection section 20 according to a setup beforehand defined by the command from the input section 10, or the proper storage section (S101). In addition, the setup is unnecessary when the selection section 20 is omitted. In a setup of the 1st speech recognition section 5, the threshold of a kanji incorrect recognition rate, the language database to be used are set up, for example. In a setup of the image delay section 2, a setup or selection of the time delay of a speaker image is performed, for example. Furthermore, the processing section 11 sets up the layout setting section 8 according to a setup beforehand defined by the command from the input section 10, or the proper storage section (S103). the [which is displayed on the alphabetic character graphic display section 9 in a setup of the layout setting section 8] -- the display condition and layout of 1 visible language data (this example a Japanese character string or a foreign language character string) and delay image data are set up. About the number of presentation character strings, a presentation graphic size, a font and a color, the display position of a character string, and delay image data, the magnitude of a speaker image, a display position, etc. are set [data / visible language] up suitably, respectively, for example.
[0029] A camera 1 inputs Speaker's A image (S105). The image delay section 2 performs a proper image processing for the image inputted into the camera 1 delay and if needed according to a setup and control by the processing section 11, and outputs delay image data (S107).
[0030] The 1st voice input section 3 inputs the voice by the 1st repetition person B or the 2nd repetition person C (S111). the 1st language (this example Japanese or a foreign language) by the 1st repetition person B by whom the 1st speech recognition section 5 was inputted into the 1st voice input section 3 according to a setup and control by the processing section 11, or the 2nd repetition person C -- recognizing -- the -- it changes into 1 visible language data (this example a Japanese character string or a foreign language character string) (S113). the [furthermore, / to which the character representation section 7 was outputted from the 1st

speech recognition section 5 if needed] -- 1 visible language data are displayed (S115).
[0031] a setup and control according [the layout setting section 8] to the processing section 11 -- following -- the [from the 1st speech recognition section 5] -- delay image data are inputted from 1 visible language data and the image delay section 2, the display layout of these data is set up, the need is accepted, and a display image is generated and outputted by the proper image processing (S121). the alphabetic character graphic display section 9 -- the output from the layout setting section 8 -- following -- the -- 1 visible language data and the image delay section 2 are displayed suitably (S123).
[0032] The processing section 11 performs return processing to step S101, when there is setting modification (S125). Moreover, when there is no setting modification, when there is no speaker A modification, it moves to the processing after step S103, and on the other hand, when there is speaker A modification, the processing section 11 can end processing (S127), and can perform processing anew.
[0033] 3. Gestalt drawing 5 of the 3rd operation is the outline block diagram of the gestalt of operation of the 3rd of a voice image inverter. The 3rd person, such as a repetition person, changes a speaker's spoken language information into alphabetic character language information, and the gestalt of this operation is showing those language information and the non-language information by the speaker through a telecommunication circuit, and assists communication between different language systems of a speaker and a user. The gestalt of this operation supports the communication in the meeting, the meeting, the lecture, lesson, education, etc. in which two or more language, such as an international congress, a meeting between many countries, and a meeting between two nations, participates especially like the gestalt of the 1st operation. The voice image inverter of the gestalt of this operation is equipped with the equipment 100 for speakers, the equipment 200 for translators, the equipments 300 and 400 for the 1st and 2nd repetition persons, the 1st and 2nd recognition equipments 500 and 600, a display 700, and a telecommunication circuit 800. The equipment 100 for speakers is equipped with a microphone a camera 1 and if needed. The equipment 200 for translators is equipped with an earphone and a microphone. The equipments 300 and 400 for the 1st and 2nd repetition persons are equipped with the 1st and 2nd voice input sections 3 and 4 and an earphone, respectively. The 1st and 2nd recognition equipments 500 and 600 are equipped with the 1st and 2nd speech recognition sections 5 and 6, input section 10-b and 10-c, processing section 11-b, and 11-c, respectively. A display 700 is equipped with the image delay section 2, the character representation section 7, the layout setting section 8, the alphabetic character graphic display section 9, input section 10-c, and processing section 11-c. Moreover, the configuration shown by drawing bullet round mark - is a telecommunication circuit 800, and means that the interface in each equipment 100-700 with which various telecommunication circuits, such as the Internet, LAN, wireless LAN, a cellular phone, and PDA, and a telecommunication circuit are inputted and outputted is established. Each of the equipment 100 for speakers, the equipment 200 for translators, the equipments 300 and 400 for the 1st and 2nd repetition persons, the 1st and 2nd recognition equipments 500 and 600, and a display 700 is suitably connected by such telecommunication circuit 800 if needed, and voice and/or a video signal communicate. You may make it connect by the direct cable or wireless, without minding the telecommunication circuit 800 of either of the drawings. Therefore, by using the telecommunication circuit 800 which has a telecommunication circuit and an interface, the display 700 installed in Speaker A, Translator D, the 1st and 2nd repetition persons B and C, the 1st and 2nd recognition equipments 500 and 600, the hall, etc. may exist anywhere, and can be arranged suitably. The configuration and actuation of a camera 1, the image delay section 2, the 1st and 2nd voice input sections 3 and 4, the 1st speech recognition section 5, the character representation section 7, the layout setting section 8, the alphabetic character graphic display section 9, and the input section 10 (-a, b, c) processing section 11 (-a, b, c) are the same as that of it of the same sign of the gestalt of the 1st operation. However, input section 10-a performs the data input directions to various setup, a proper database, memory of each part of the image delay section 2 and layout setting section 8 grade, etc. Processing section-a is a small computer and controls each part of the image delay section 2, input section 10-a, -b and 10-c, and layout setting section 8 grade. Moreover, input

section 10-b and 10-c perform the data input directions to various setup, a proper database, memory of the 1st and 2nd speech recognition sections 5 and 6, etc. Processing section 11-b and 11-c are small computers, and control the 1st and 2nd speech recognition section 5 and each part of 6 grades. Moreover, the flow chart of voice transform processing of the gestalt of the 3rd operation is the same as that of the gestalt of the 1st operation, and as mentioned above, it operates.

[0034] 4. Gestalt drawing 6 of the 4th operation is the outline block diagram of the gestalt of operation of the 4th of a voice image inverter. The 3rd person, such as a repetition person, changes a speaker's spoken language information into alphabetic character language information, and the gestalt of this operation is showing those language information and the non-language information by the speaker through a telecommunication circuit, and assists communication between different language systems of a speaker and a user. The gestalt of this operation supports the communication in the meeting, the meeting, the lecture, lesson, education, etc. in which two or more language, such as an international congress, a meeting between many countries, and a meeting between two nations, participates especially like the gestalt of the 3rd operation. The voice image inverter of the gestalt of this operation is equipped with the equipment 100 for speakers, the equipment 200 for translators, the equipments 300 and 400 for the 1st and 2nd repetition persons, the 1st recognition equipment 500, a display 700, and a telecommunication circuit 800.

[0035] Although it differs in that the 2nd recognition equipment 600 which contains the 2nd speech recognition section as compared with the gestalt of the 3rd operation was omitted, and the 1st recognition equipment 500 was further equipped with the selection section 20, other configurations and actuation are the same. The configuration and actuation of the selection section 20 are the same as that of the gestalt of the 2nd operation. In addition, the 2nd voice input section and the selection section 20 may be omitted further if needed. Moreover, the flow chart of voice transform processing of the gestalt of the 4th operation is the same as that of the gestalt of the 3rd operation, and as mentioned above, it operates.

[0036] 5. When a voice recognition unit inputs into this voice recognition unit the voice as which the repetition person repeated Speaker's A voice using a beforehand registered repetition person's voice database, voice conversion is carried out and what kind of speaker A is made for a high recognition rate to be acquired as mentioned above with the gestalt of the conclusion book operation. When Speaker A is Translator D, and a repetition person repeats Translator's D voice, a foreign language can be translated into Japanese by the high recognition rate. On the contrary, in the case of the voice about which it spoke in Japanese, when Translator D translates into a foreign language and repeats the voice in the foreign language, Japanese can be translated into a foreign language by the high recognition rate. Similarly, since the character representation also of a questioner's voice can be carried out, bidirectional meeting exchange is realizable. Therefore, the gestalt of this operation can be used also as communication exchange not only in a domestic meeting but an international congress.

[0037] Moreover, according to the gestalt of this operation, Speaker's A image was also incorporated, the approach of displaying together with the character string of a recognition result by a certain time delay is taken, and the image of sign language etc. can be used for a motion of Speaker's A lip, and an expression pan as a key of a speech understanding. According to a hearing-impaired person's speech reading capacity, the image time delay by the image delay section 2 can be changed now. Therefore, for the hearing-impaired person who became skilled in the speech reading which reads a motion of a lip, 5% of error of speech recognition is restorable by speech reading.

[0038] The alphabetic character image conversion approach of this invention, or an alphabetic character image inverter and a system can be offered by computers, such as a server which includes the program product in which loading is possible, and its program in the internal memory of a computer including the record medium which recorded the alphabetic character image conversion program for making a computer perform each of that procedure, and the alphabetic character image conversion program, and in which computer reading is possible, and an alphabetic character image conversion program, etc.

[0039]
[Effect of the Invention] While according to this invention a repetition person changes the voice of an unspecified speaker into self voice and changes it into an alphabetic character through a voice recognition unit as mentioned above, a voice image inverter for a hearing-impaired person etc. to make easy to understand what the speaker spoke about and an approach, and a voice image conversion program can be offered by delaying images, such as a speaker's expression, and displaying on a screen etc. with an alphabetic character.

[0040] Moreover, according to this invention, in the meeting of an international congress which a hearing-impaired person attends, a bilateral meeting between many countries, etc., etc., a repetition person can repeat the voice of a lecturer or a translator, it can input into a voice recognition unit, and the voice image inverter for the meeting exchange which displayed the character string which it is as a result on the screen with a lecturer's image and an approach, and a voice image conversion program can be offered. Furthermore, according to this invention, a user can be provided with the information by which transmitted voice to the repetition person and literation was carried out from exchange of the meeting where the interpreter of the international congress performed using different-species language and the immediate printing (information compensation) of a meeting, a hearing-impaired person, etc. participate, or a lesson, and a telephone. Moreover, according to this invention, the voice image inverter for assisting communication between different language systems of a speaker and a user and an approach, and a voice image conversion program can be offered. Moreover, wherever a user may be in, it can make it possible to use this system by adding a means to transmit a speaker's voice and image to the translator, the repetition person, and those [correction] who are present in a remote place or the home ground by the telecommunication circuit which communicates using telecommunication circuits, such as the Internet, further according to this invention. According to this invention, working is supportable when the trouble back tone of difficult being home of going out turns into a repetition person further, that the intervening repetition person and a translator use as home business, and.

[Translation done.]

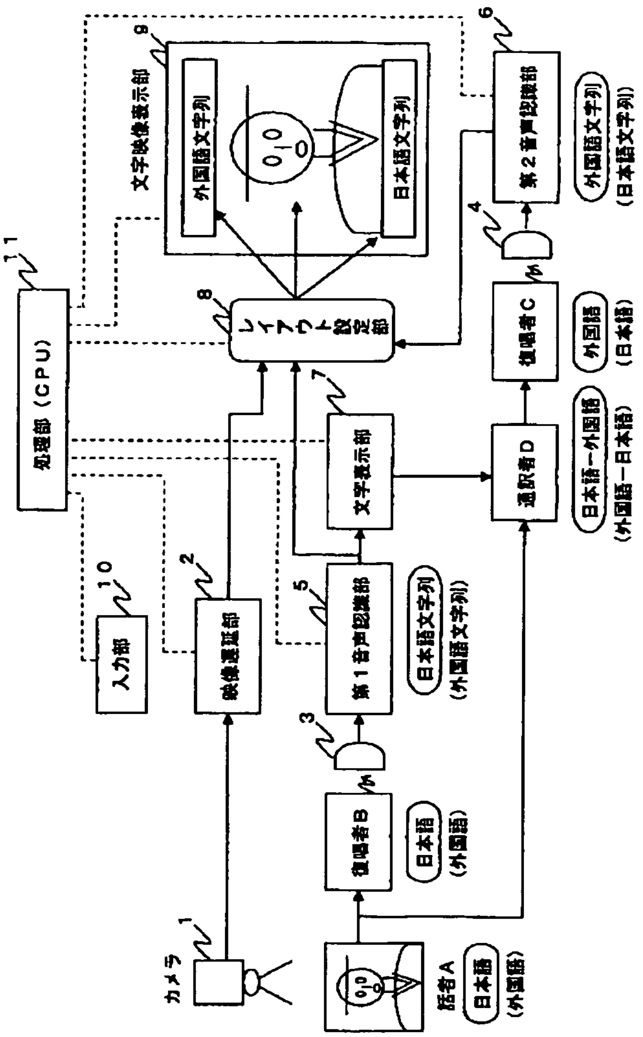
* NOTICES *

JPO and NCIPI are not responsible for any damages caused by the use of this translation.

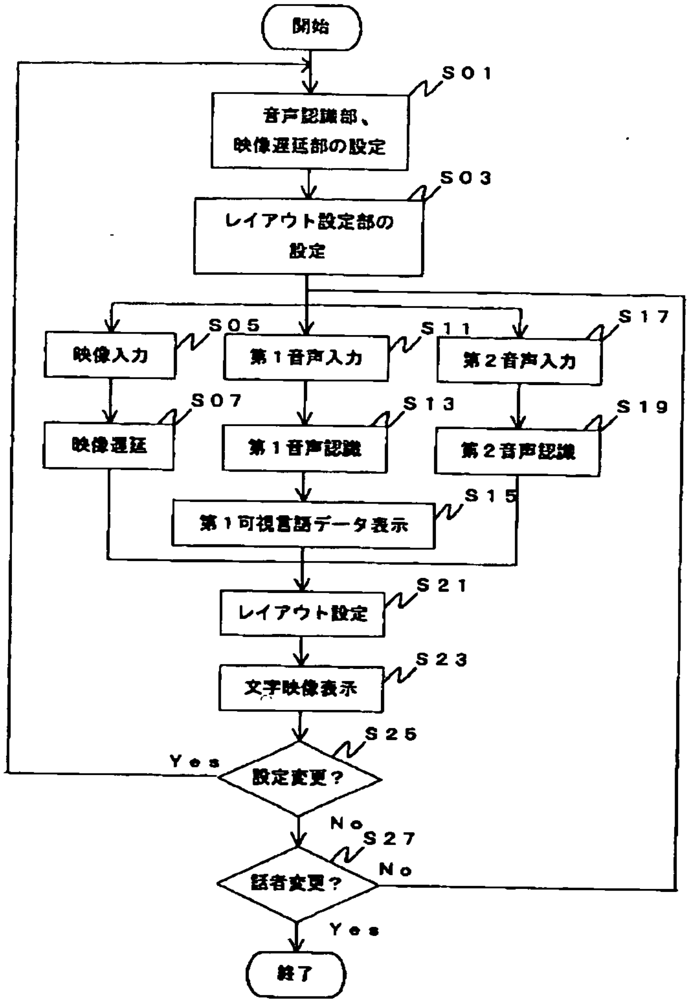
1.This document has been translated by computer. So the translation may not reflect the original precisely.
2.**** shows the word which can not be translated.
3.In the drawings, any words are not translated.

DRAWINGS

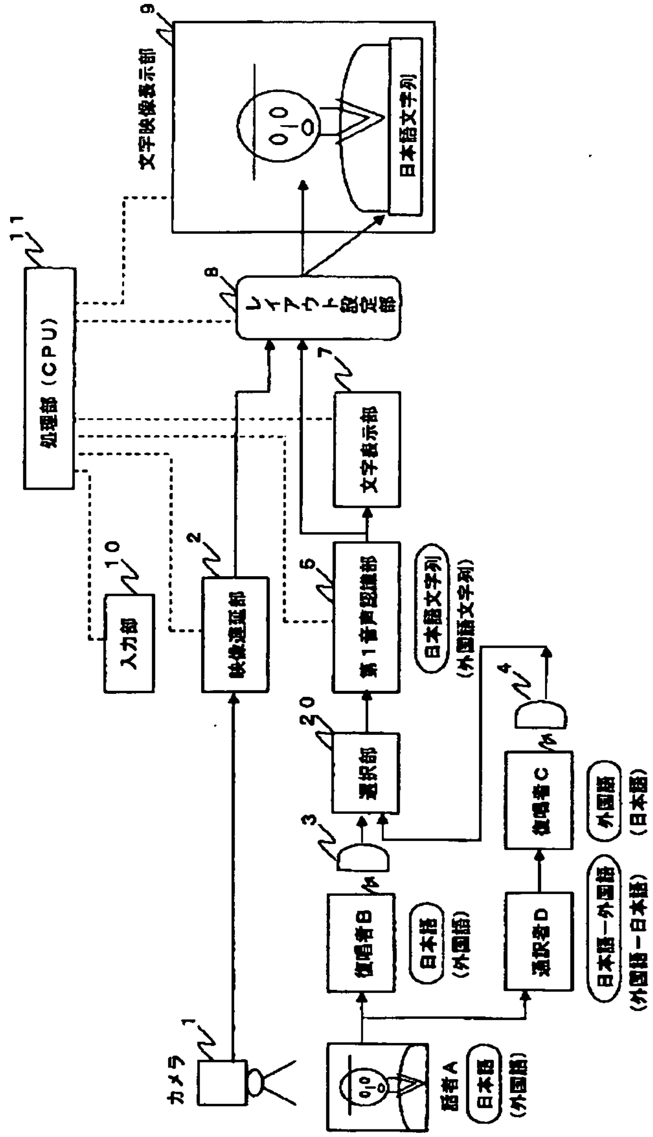
[Drawing 1]


カメラ
1
話者A
日本語
(外国語)
復唱者B
日本語
(外国語)
3
映像選択部
2
入力部
10
処理部（CPU）
11
第1音声認識部
5
日本語文字列
(外国語文字列)
文字表示部
7
通訳者D
日本語－外国語
(外国語－日本語)
復唱者C
外国語
(日本語)
4
第2音声認識部
6
外国語文字列
(日本語文字列)
レイアウト設定部
8
文字映像表示部
9
外国語文字列
日本語文字列

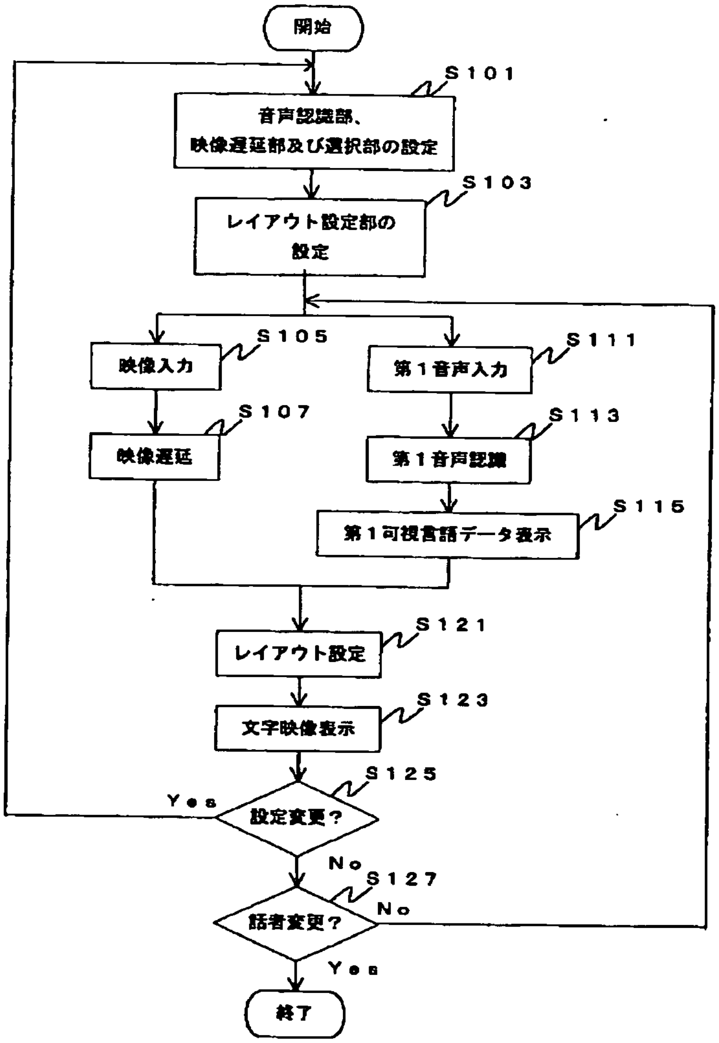
[Drawing 2]


開始
S01
音声認識部、
映像遅延部の設定
S03
レイアウト設定部の
設定
S05
映像入力
S11
第１音声入力
S17
第２音声入力
S07
映像遅延
S13
第１音声認識
S19
第２音声認識
S15
第１可視言語データ表示
S21
レイアウト設定
S23
文字映像表示
S25
Yes
設定変更？
No
S27
No
話者変更？
Yes
終了

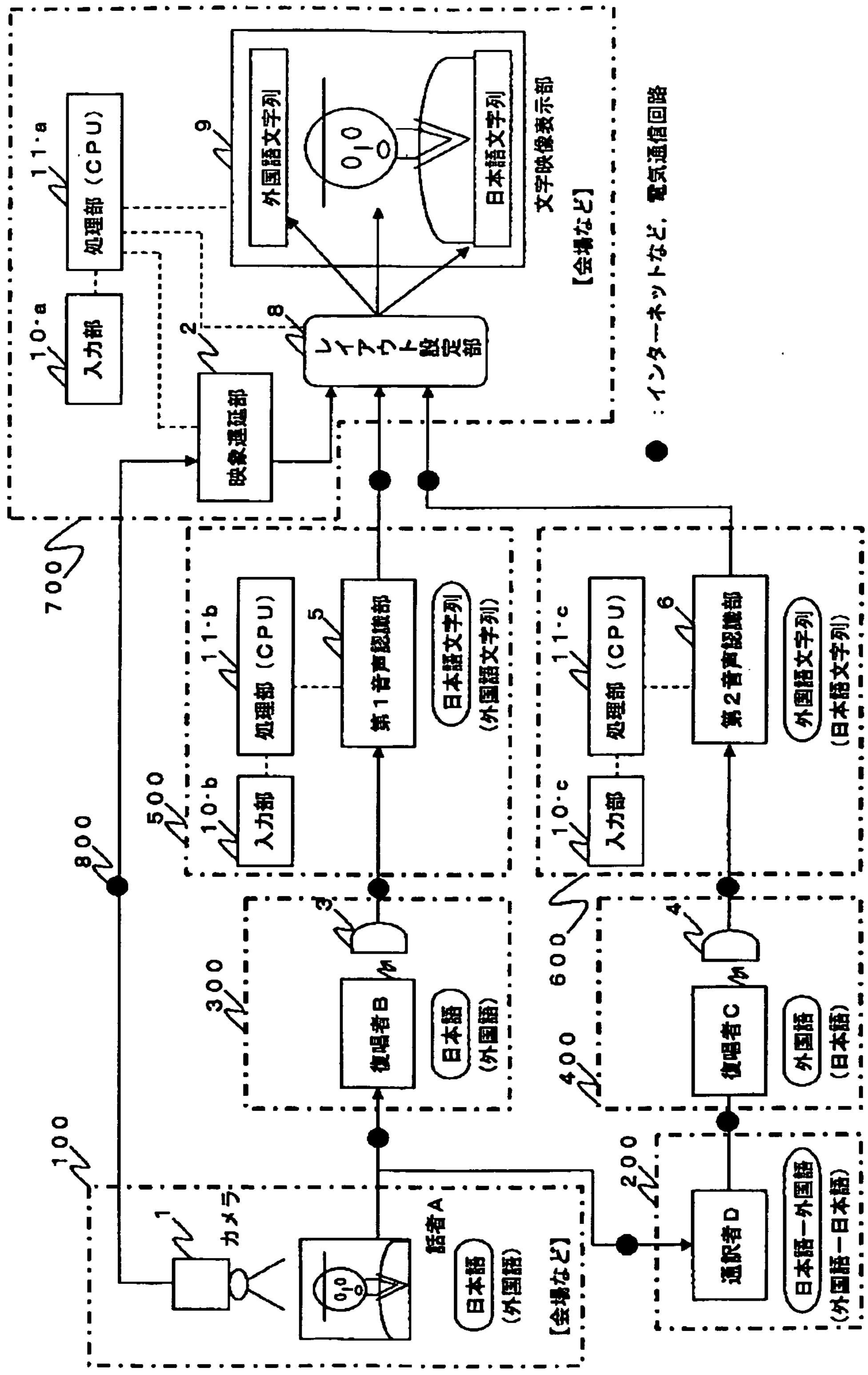
[Drawing 3]


カメラ
1
話者A
日本語
(外国語)
復唱者B
日本語
(外国語)
通訳者D
日本語ー外国語
(外国語ー日本語)
3
選択部
20
復唱者C
外国語
(日本語)
4
第１音声認識部
5
日本語文字列
(外国語文字列)
文字表示部
7
映像遅延部
2
入力部
10
処理部（CPU）
11
レイアウト設定部
8
文字映像表示部
9
日本語文字列

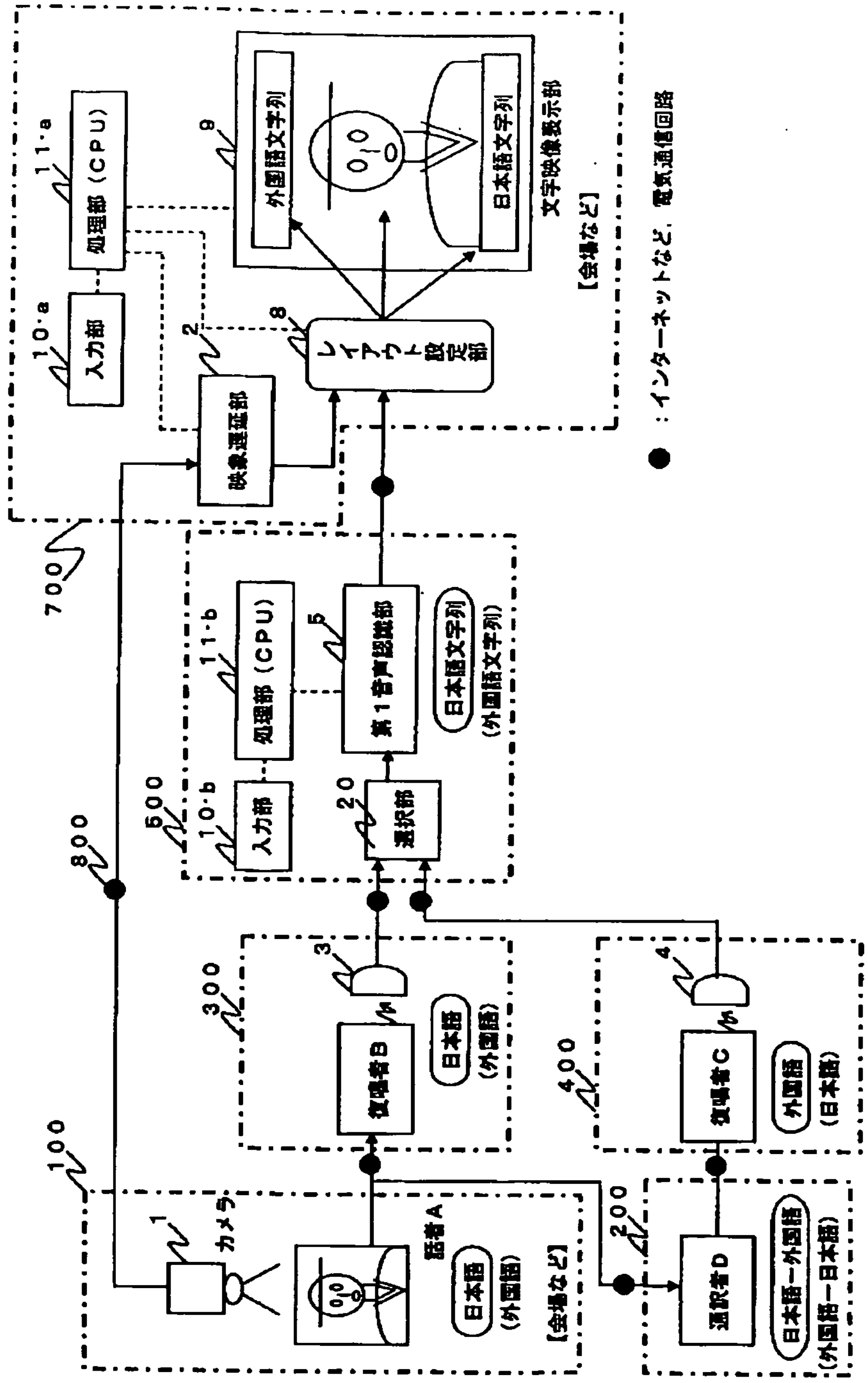
[Drawing 4]


開始
S101
音声認識部、映像遅延部及び選択部の設定
S103
レイアウト設定部の設定
S105
映像入力
S107
映像遅延
S111
第１音声入力
S113
第１音声認識
S115
第１可視言語データ表示
S121
レイアウト設定
S123
文字映像表示
S125
設定変更？
Yes
No
S127
話者変更？
No
Yes
終了

[Drawing 5]


処理部（CPU）
11-a
入力部
10-a
映像遅延部
2
レイアウト設定部
8
9
外国語文字列
日本語文字列
文字映像表示部
【会場など】
●：インターネットなど，電気通信回路
700
500
入力部
10-b
処理部（CPU）
11-b
第1音声認識部
5
日本語文字列
（外国語文字列）
600
入力部
10-c
処理部（CPU）
11-c
第2音声認識部
6
外国語文字列
（日本語文字列）
800
300
復唱者B
3
日本語
（外国語）
400
復唱者C
4
外国語
（日本語）
100
カメラ
1
話者A
日本語
（外国語）
【会場など】
200
通訳者D
日本語→外国語
（外国語→日本語）

[Drawing 6]


処理部（CPU）
11·a
入力部
10·a
映像遅延部
2
レイアウト設定部
8
外国語文字列
日本語文字列
文字映像表示部
9
【会場など】
700
●：インターネットなど、電気通信回路
処理部（CPU）
11·b
入力部
10·b
第１音声認識部
5
選択部
20
日本語文字列
（外国語文字列）
500
800
復唱者B
3
日本語
（外国語）
300
復唱者C
4
外国語
（日本語）
400
カメラ
1
話者A
日本語
（外国語）
【会場など】
100
通訳者D
日本語ー外国語
（外国語ー日本語）
200

[Translation done.]

This Page is Inserted by IFW Indexing and Scanning Operations and is not part of the Official Record

BEST AVAILABLE IMAGES

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images include but are not limited to the items checked:

☐ **BLACK BORDERS**

☐ **IMAGE CUT OFF AT TOP, BOTTOM OR SIDES**

☑ **FADED TEXT OR DRAWING**

☑ **BLURRED OR ILLEGIBLE TEXT OR DRAWING**

☐ **SKEWED/SLANTED IMAGES**

☐ **COLOR OR BLACK AND WHITE PHOTOGRAPHS**

☐ **GRAY SCALE DOCUMENTS**

☑ **LINES OR MARKS ON ORIGINAL DOCUMENT**

☑ **REFERENCE(S) OR EXHIBIT(S) SUBMITTED ARE POOR QUALITY**

☐ **OTHER:** ____________________

IMAGES ARE BEST AVAILABLE COPY.
As rescanning these documents will not correct the image problems checked, please do not report these problems to the IFW Image Problem Mailbox.